

## Japanese Idiom variants in corpus data and phraseological dictionaries

|                              |   |
|------------------------------|---|
| 著者                           | Ishida Priscilla  |
| journal or publication title | 筑波応用言語学研究   |
| number                       | 21  |
| page range                   | 16-30   |
| year                         | 2014-12-20  |
| URL                          | <a href="http://hdl.handle.net/2241/00123584">http://hdl.handle.net/2241/00123584</a> |

# Japanese idiom variants in corpus data and phraseological dictionaries

Priscilla Ishida

Keywords: idioms, variants, corpus, phraseological dictionaries

## 1. Introduction

Idioms are multi-word expressions characterized by strong bonding between their individual constituents.<sup>1</sup> According to traditional views, this means it is not possible to replace individual words with synonyms or antonyms (1), or to delete them (2).

(1) *saji*/\**spūn o nageru* (lit. throw the spoon (Jap.)/\*spoon (Eng. loanword), ‘give up’), *abura o uru*/\**kau* (lit. sell/\*buy oil, ‘loaf; waste time’)

(2) *kumo no ko*/\**kumo o chirasu yō* (lit. as if one scatters spider babies/\*spiders, ‘[run off] in all directions’)

However, newspapers, books, internet blogs, and other Japanese texts yield evidence that some idioms do allow the replacement (3) or deletion (4) of constituents.

(3) *Shikashi, gakkō kankeisha wa ijime ga atta ka dō ka wa wakaranai to kuchi o nigosu*. (‘However, the parties concerned within the school prevaricate (lit. muddy their mouths), saying that they don’t know whether or not there was any bullying.’) [book]

cf. *kotoba o nigosu* (lit. muddy one’s words, ‘speak evasively; prevaricate’)

(4) *Me kara uroko no rekishi kōza*. (‘It’s an eye-opening (lit. scales-from-the-eyes) history course.’) [blog]

cf. *me kara uroko ga ochiru* (lit. the scales fall from one’s eyes, ‘be awakened to the truth’)

This evidence is consistent with studies of idioms in languages such as English (Moon 1998; Philip 2008) and German (Itō 1990), which show that while some idioms are quite frozen, others have pairs or sets of variant forms with synonymous or antonymous meanings. This paper takes the view that pairs of expressions such as *kotoba o nigosu/kuchi o nigosu* (3) and *me kara uroko ga ochiru/me kara uroko* (4) are “idiom variants,” which are defined as two or more idiomatic expressions with shared lexis and meaning that have the status of relatively stable units of the Japanese lexicon (Ishida 1998; cf. Itō 1990; Moon 1998). The aim of this paper is to show what corpus

---

<sup>1</sup> Miyaji (1982a: 238) defines idioms as “strings of two or more words that have comparatively strong internal bonding and fixed, holistic meanings” (*tango no futatsu ijō no renketsutai de atte, sono musubitsuki ga hikaku-teki kataku, zentai de kimatta imi o motsu kotoba*). This paper will adopt Miyaji’s (1982a) definition and deal with expressions that fall within its scope.

data reveals about idiom variants and to investigate how well the phenomenon of variability is represented in phraseological dictionaries.

## 2. Past views of fixedness and variability

“Fixedness” or “stability” is widely considered to be a fundamental characteristic of idioms. Past research defines Japanese idioms as strings of words that are habitually used together and resistant to the replacement, addition, and deletion of lexical constituents (Miyaji 1982a; Muraki 1985, 1991; Kunihiro 1985; Itō 1997).

However, there is evidence that many idioms have one or more variant forms that are strongly institutionalized. Miyaji (1982b) and Morita (1985, 1994) point out that some transitive verb phrase idioms have intransitive counterparts, and vice versa (e.g. *yaridama ni ageru* lit. raise on the tip of a spear, ‘make an example of’/*yaridama ni agaru* lit. rise on the tip of a spear, ‘be made an example of’). Miyaji (1985) shows that some verb and adjective phrase idioms have corresponding compound forms (e.g. *me ga sameru* lit. one’s eyes awake, ‘wake up to; realize’/*mezameru* ‘wake up to; realize’). In addition, Morita (1985, 1994) notes that some idioms have synonymous or antonymous variants created by the replacement of one of their content words (e.g. *ude ga sagaru/ochiru* lit. one’s arm goes down/falls, ‘one’s ability or performance declines’; *koshi ga tsuyoi/yowai* lit. one’s hips are strong/weak, ‘resolute/weak-kneed’).

Ishida (1998) reports the results of a systematic analysis and classification of Japanese idiom variants, based on data collected from post-war novels, and identifies two principal types of variation. The first is structural variation, which involves the addition or deletion of constituents and a change in the internal structure and/or grammatical function of the idiom. The second is lexical variation, which arises from the alternation of a noun, verb, or adjective constituent.

Satō’s (2007a, 2007b) analysis of the idiom headwords in five Japanese dictionaries found inconsistencies in the representation of the lexical structure and/or phonological form of some expressions. Satō attributes some of these inconsistencies to idiom variants (which are exemplified but not discussed in detail), and others to differences in individual lexicographers’ intuitions of the degree of “standardization” of idiom forms. He also suggests that the lack of a “standard” idiom dictionary and the lack of set procedures for establishing headword forms are problematic for dictionary editors.

It is clear that past research on Japanese idioms has explored, to some extent, the phenomenon of idiom variability. However, until now, scholars have relied on intuition, manually collected data, and/or genre-specific electronic databases. This paper takes the view that corpus data can be used to gain a clearer picture of the phenomenon of variability, including types of variants, non-intuitive forms, and frequency of use. Sections 3 and 4 of this paper will report the methodology and results of a corpus analysis of idiom variants, and Section 5 will investigate how well evidence from corpus data is reflected in monolingual Japanese phraseological dictionaries.

### 3. Extracting idiom variants from corpus data

#### 3.1 Target idioms, tools, and procedures

In order to identify variant forms, this researcher first selected a set of 100 commonly-used idioms dealt with in past research and then collected potential candidates for variants from the Balanced Corpus of Contemporary Written Japanese (BCCWJ) and an opportunistic web corpus (Tsukuba Web Corpus, TWC). In order to find all possible forms of the target idioms, queries were created from the smallest possible number of constituents (Moon 1998; Fellbaum et al. 2006; Philip 2008), using the online search tools described below.

- 1) **NINJAL-LWP for BCCWJ (NLB) ver. 1.20:** This is an online lexical profiling tool developed jointly by the National Institute for Japanese Language and Linguistics (NINJAL) and the Lago Institute of Language. It uses the 2011 DVD version of the BCCWJ, which includes 11 text types and approximately 105 million words. The purpose of using this tool was to identify potential variant forms by extracting all instances of a key word in the target idiom and filtering the results yielded for phrase structures and collocates.
- 2) **NINJAL-LWP for TWC (NLT) ver. 1.10:** NLT is an online lexical profiling tool for the Tsukuba Web Corpus (TWC), which is a corpus of 1.1 billion words compiled from Japanese websites in January, 2012 by researchers at the University of Tsukuba. It has the same interface as NLB (see above). The purpose of using NLT was to confirm potential variant candidates previously identified in the BCCWJ, using a larger corpus.

Multiple queries were necessary for both tools. After the initial search results were obtained, irrelevant data (e.g. literal examples and duplicates) were eliminated manually. The criteria used to identify idiom variants were shared lexis and meaning and an arbitrary threshold of at least 10 occurrences in the BCCWJ. In borderline cases, native speaker judgments of familiarity/acceptability were also taken into account.<sup>2</sup>

Search results were compiled for each variant, and the alternating forms of each idiom were then analyzed and classified. For the purpose of the dictionary analysis, the target idioms were limited to a set of 56 (with a total of 150 individual forms) exemplifying the five types of variants shown below. This paper will compare the results of the corpus analysis and dictionary analysis of synonymous and antonymous variants.

- |                                     |                           |
|-------------------------------------|---------------------------|
| 1) synonymous variants              | 4) idioms/compounds       |
| 2) antonymous variants              | 5) truncations/expansions |
| 3) transitive/intransitive variants |                           |

---

<sup>2</sup> The BCCWJ yields only 8 instances of *kenka o kau* (lit. buy a fight, ‘accept a challenge’, Table 4). However, since native speakers judge this expression to be acceptable and familiar, it is viewed in this paper to be a stable variant.

## 3.2 Problems of identification and description

### 3.2.1 The question of “canonical” forms

Idiom dictionaries tend to present idioms as “syntactically, morphologically, and lexically fixed canonical forms” (Fellbaum et al. 2006: 350), and language users tend to believe that the dictionary look-up form of an idiom is its “standard” form. One possible view is that idioms have a canonical form and variants are deviations from this form. Judgments regarding which is the canonical form and which is (are) the variant(s) are usually made on the basis of frequency, derivational relationships, or provenance. However, in some cases multiple forms of an idiom are used frequently, as in (5), and language users may judge multiple forms to be “standard” in spite of differences in frequency, as in both (5) and (6).

(5) *hara ga tatsu* (lit. one’s belly rises, ‘get angry’) BCCWJ 766

*hara o tateru* (lit. raise one’s belly, ‘get angry’) BCCWJ 570

(6) *shūshifu o utsu* (lit. strike [typewrite] a period, ‘put an end to s.t.’) BCCWJ 164

*piriodo o utsu* (lit. strike [typewrite] a period, ‘put an end to s.t.’) BCCWJ 42

Another possible view is that pairs such as those shown in (5) and (6) are one idiom with alternate forms (Moon 1998; Ishida 1998). In this view, *hara ga tatsu* is a variant of *hara o tateru* (5) and vice versa. This view places emphasis on the mutual relationships between alternating forms and also provides a framework flexible enough to handle sets or clusters of forms with variable lexis and structure (Moon 1998; Philip 2008). This paper takes the second view—that alternating idiom forms are variants of each other—and does not distinguish between “canonical” and “non-canonical” forms.

### 3.2.2 Lexico-structural variants vs. independent idioms

It is sometimes difficult to distinguish between pairs/sets of idiom variants, on one hand, and different idioms that happen to belong to the same semantic field, on the other. For instance, the expressions in (7) all share the elements *mimi* ‘ear’) and *ni* (DAT ‘into’) and convey the meaning ‘hear s.t. unintentionally’, so at first glance they may appear to constitute a set of variants.

(7a) *mimi ni hasamu* (lit. insert s.t. in one’s ear)

(7b) *komimi ni hasamu* (lit. insert s.t. in one’s little-ear)

(7c) *mimi ni suru* (lit. do s.t. into one’s ear)

(7d) *mimi ni hairu* (lit. s.t. enters one’s ear)

However, a closer look at the lexis and meaning of each idiom shows that a more finely-tuned distinction is called for. *Mimi ni hasamu* (7a) and *komimi ni hasamu* (7b) share the elements *mimi* ‘ear’ and *hasamu* ‘insert’, as well as the same syntactic structure and meaning; the element *ko* ‘small’ preceding *mimi* ‘ear’ simply emphasizes the lack of intentionality of the action. On the other hand, *mimi ni suru* (7c) and *mimi ni hairu* (7d) share less lexical material with each other and with the first two idioms; there

are also differences in meaning and usage (Miyaji 1982a).<sup>3</sup> Thus, although all four expressions share lexis and the general meaning of ‘hear unintentionally,’ only *mimi ni hasamu/komimi ni hasamu* (7a/7b) can be classed as idiom variants. The implication is that, in order to distinguish between idiom variants and independent idioms with related meaning (7c/7d), it is necessary to compare not only lexis but also the meaning and usage of each expression.

## 4. Idiom variants in corpus data

### 4.1 Synonymous variants

Results of the corpus analysis indicate that some idioms allow the alternation of a content word, with little or no change to the internal structure and meaning of the idiom. Table 1 shows synonymous variants produced by alternation of the noun constituent of an idiom. In some cases, the alternation includes a noun of Japanese origin on one hand, and a loanword on the other. For example, *shūshifu* ‘period’ and *hadome* ‘brake; skid’ are both nouns of Japanese origin, and *piriodo* ‘period’ and *burēki* ‘brake(s)’ are both borrowings from English. With respect to the pair *shūshifu o utsu* and *piriodo o utsu*, the latter expression is held to be the original form, with *shūshifu o utsu* appearing after *shūshifu* was introduced as a Japanese equivalent to the loanword *piriodo* sometime during the Showa Era (Miyaji 1982a: 251). In present-day Japanese *shūshifu o utsu* is used more frequently (Table 1), but both expressions are common and familiar.

Table 1 Synonymous variants (noun alternation)

| idioms/variants          | meaning (literal/idiomatic)                              | BCCWJ*     | TWC          |
|--------------------------|--|------------|--------------|
| <i>mi ni shimiru</i>     | lit. permeate one’s flesh, ‘touch/sting s.o. deeply’     | 212 (2.02) | 2,422 (2.13) |
| <i>honemi ni shimiru</i> | lit. permeate one’s bones and flesh                      | 23 (0.22)  | 163 (0.14)   |
| <i>shūshifu o utsu</i>   | lit. strike [typewrite] a period (Jap.), ‘put an end to’ | 164 (1.56) | 1,397 (1.23) |
| <i>piriodo o utsu</i>    | lit. strike [typewrite] a period (Eng. loanword)         | 42 (0.40)  | 340 (0.30)   |
| <i>hadome o kakeru</i>   | lit. put on the brake/skid (Jap.), ‘slow down/stop s.t.’ | 96 (0.92)  | 1,229 (1.08) |
| <i>burēki o kakeru</i>   | lit. put on the brake(s) (Eng. loanword)                 | 61 (0.58)  | 614 (0.54)   |
| <i>kotoba o nigosu</i>   | lit. muddy one’s words, ‘speak evasively/ambiguously’    | 80 (0.76)  | 249 (0.22)   |
| <i>kuchi o nigosu</i>    | lit. muddy one’s mouth                                   | 16 (0.15)  | 59 (0.05)    |

\*Figures show raw numbers of instances in each corpus and (in parentheses) frequency per million words.

With respect to the pair *kotoba o nigosu/kuchi o nigosu* (Table 1), *kotoba o nigosu* is popularly considered to be the “standard” form, but the BCCWJ also yields 16

<sup>3</sup> According to Miyaji (1982a: 100, 194), *mimi ni suru* (7c) means to hear sounds, voices, and so on, while *(ko)mimi ni hasamu* (7a/7b) means to hear other people’s conversations or a cohesive piece of information such as news, a rumour, etc. Also, the argument structure and meaning of *mimi ni hairu* (7d) parallel those of the verb *kikoeru* ‘reach one’s ears’, while *mimi ni suru* (7c) is closer to *kiku* ‘hear’; thus *mimi ni hairu* highlights the receptivity of the agent.

occurrences of the variant *kuchi o nigosu*. This is consistent with the results of a survey on language use conducted by the Japanese Agency for Cultural Affairs in 2005, which showed that while 66.9 % of respondents tend to use *kotoba o nigosu*, 27.6 % use *kuchi o nigosu* and 3.1% use both.<sup>4</sup>

**Table 2** Synonymous variants (verb alternation)

| idioms/variants           | meaning (literal/idiomatic)  | BCCWJ     | TWC        |
|---------------------------|--|-----------|------------|
| <i>kao-iro o ukagau</i>   | lit. glance at s.o.'s face-colour, 'judge s.o.'s feelings/ thoughts from his/her expression' | 98 (0.94) | 778 (0.68) |
| <i>kao-iro o miru</i>     | lit. look at s.o.'s face-colour  | 37 (0.35) | 235 (0.21) |
| <i>kao-iro o yomu</i>     | lit. read s.o.'s face-colour   | 10 (0.10) | 17 (0.01)  |
| <i>kageguchi o tataku</i> | lit. beat the shadow-mouth [=malicious gossip], 'backbite'                                   | 59 (0.56) | 257 (0.23) |
| <i>kageguchi o iu</i>     | lit. speak/say the shadow-mouth  | 20 (0.19) | 197 (0.17) |
| <i>kageguchi o kiku</i>   | lit. make use of the shadow-mouth  | 14 (0.13) | 16 (0.01)  |

There are also pairs or sets of variants with alternating verbs (Table 2). In some cases, the verbs themselves have broadly related meanings, as do (*kao-iro o*) *ukagau* 'glance at', *miru* 'look', and *yomu* 'read'. Both the BCCWJ and the TWC add a few instances of other verbs to this paradigm, including (*kao-iro o*) *saguru* 'search' (BCCWJ 5) and *tashikameru* 'check; see [if]' (TWC 3). Because of their low frequency, the latter expressions do not have the status of variants as defined in this paper. However, they do provide further evidence of the productivity of the set *kao-iro o ukagau/miru/yomu*.

On the other hand, in the case of *kageguchi o tataku/iu/kiku*, the alternating idiom phrases are synonymous in spite of the fact that the meanings of the constituent verbs are unrelated (*tataku* 'hit; beat', *iu* 'speak; say' and *kiku* 'make use of').

In addition to the synonymous pairs and sets shown in Tables 1 and 2, there are also cases of rather free alternation among a variety of noun or verb constituents (Table 3).

**Table 3** Variants with empty slots

| idioms/variants                          | meaning (literal/idiomatic)                                    | BCCWJ         | TWC          |
|--|--|---------------|--------------|
| <i>yumizu no yō ni (gotoku)...V</i>      | V...like hot and cold water, '[waste/drain away]...like water' | 43 (0.41)     | 371 (0.33)   |
| <i>yumizu no yō ni (gotoku)...tsukau</i> | use...like hot and cold water, 'squander/waste...like water'   | [31 (0.30)] * | [187 (0.16)] |
| <i>chi no nijimu yō na N</i>             | N like blood oozing, 'strenuous/backbreaking N'                | 28 (0.27)     | 255 (0.22)   |
| <i>chi no nijimu yō na do'ryoku</i>      | effort like blood oozing, 'strenuous/backbreaking effort'      | [19 (0.18)]   | [157 (0.14)] |

\*[31] indicates that out of 43 instances of '*yumizu no yō ni...V*', 31 included the verb *tsukau* 'use'.

<sup>4</sup> Details are available on the website of the Japanese Agency for Cultural Affairs (in Japanese only). <[http://www.bunka.go.jp/kokugo\\_nihongo/yoronchousa/h17/kekka.html](http://www.bunka.go.jp/kokugo_nihongo/yoronchousa/h17/kekka.html)> Retrieved 28.02.2014.

There is a strong tendency for *yumizu no yō ni* (lit. like hot and cold water) to be followed by the verb *tsukau* ‘use’, as in (8) below. However, this expression also combines with other verbs that mean ‘use’ or ‘flow out’ in a broad sense, as in (9).

(8) *Kaiin o fuyasu tame no kōkokuhī o yumizu no yō-ni tsukatta.* (‘They squandered [lit. used like hot and cold water] the budget allocated for advertising to increase membership.’)

(9) (*yumizu no yō ni*)...*rōhi suru* ‘waste’, *shōhi suru* ‘consume; expend’, *nagareru* ‘flow’, *ryūshutsu suru* ‘flow out’, *sosogu* ‘pour’, *dete iku* ‘go out/drain away’

In the same way, the phrase *chi no nijimu yō na* (lit. like blood oozing) combines most frequently with the noun *do’ryoku* ‘effort’ (10). At the same time, this phrase is also used to modify a variety of other nouns expressing ‘an activity that demands effort or labour’ (11).

(10) *Sono hi kara, Kazuo no chi no nijimu yō na do’ryoku ga hajimari...* (‘From that day on, Kazuo’s backbreaking efforts [lit. effort(s) like blood oozing] began...’)

(11) (*chi no nijimu yō na*) *renshū* ‘practice’, *kunren* ‘training’, *shugyō* ‘training; apprenticeship’, *kurō* ‘hardship; suffering’, *kutō* ‘a tough struggle; a hard fight’

What is evident in these cases is a tension between fixedness and productivity, as well as a kind of extended patterning: a particular verb or noun tends to be preferred, but other verbs or nouns that belong to the same semantic class can be substituted rather freely. The lexical structure of the expressions in (8-11) can be represented as ‘*yumizu no yō ni*...V’ and ‘*chi no nijimu yō na* N’, respectively. The V or N in each expression is an “empty slot” filled at the time of use by a verb or noun that belongs to the particular semantic class described above (e.g. *chi no nijimu yō na kunren* ‘strenuous **training**’).<sup>5</sup>

## 4.2 Antonymous variants

Corpus data also yield cases of alternation of a verb or adjective constituent of an idiom, resulting in variants with shared lexis and structure and opposite meaning. Table 4 shows variant pairs in which an adjective or verb alternates with its usual antonym (e.g. *tsuyoi/yowai*, ‘strong/weak’; *kasu/kariru*, ‘lend/borrow’), and the idiom phrases themselves also have opposite or converse meaning (e.g. *ki ga tsuyoi/yowai* lit. one’s *ki* [spirit/mind] is strong/weak, ‘strong-willed; tough’/‘timid; fainthearted’; *te o kasu/kariru* lit. lend/borrow a hand, ‘give/get help’).

However, the antonymous element is sometimes unpredictable (Table 5), as in the case of *kuchi ga karui* (lit. one’s mouth is light, ‘unable to keep a secret’) and *kuchi ga*

<sup>5</sup> It might be argued that the V and N slots indicated above are not intra-idiomatic constituents but extra-idiomatic collocates. However, because the bonding with *tsukau* ‘use’ and *do’ryoku* ‘effort’, respectively, is relatively strong (Table 3), and because the range of alternating verbs and nouns is restricted, this paper takes the view that the V and N slots do in fact constitute parts of their respective idioms.



**Table 4 Antonymous variants (1)**

| idioms/variants             | meaning (literal/idiomatic)  | BCCWJ      | TWC          |
|-----------------------------|--|------------|--------------|
| <i>ki ga (no) tsuyoi</i>    | lit. one's <i>ki</i> [spirit/mind] is strong, 'strong-willed'            | 161 (1.54) | 1,326 (1.17) |
| <i>ki ga (no) yowai</i>     | lit. one's <i>ki</i> is weak, 'timid; fainthearted'                      | 154 (1.47) | 1,060 (0.93) |
| <i>ki ga (no) mijikai</i>   | lit. one's <i>ki</i> [spirit/mind] is short, 'short-tempered; impatient' | 55 (0.52)  | 299 (0.26)   |
| <i>ki ga (no) nagai</i>     | lit. one's <i>ki</i> is long, 'patient'                                  | 27 (0.26)  | 260 (0.23)   |
| <i>hagire ga (no) yoi</i>   | lit. the teeth-sharpness is good, 'clear and crisp; articulate'          | 98 (0.94)  | 584 (0.51)   |
| <i>hagire ga (no) warui</i> | lit. the teeth-sharpness is bad, 'evasive; inarticulate'                 | 50 (0.48)  | 368 (0.32)   |
| <i>te o kasu</i>            | lit. lend a hand, 'give help [to s.o.]'                                  | 297 (2.83) | 1,432 (1.26) |
| <i>te o kariru</i>          | lit. borrow a hand, 'get help [from s.o.]'                               | 99 (0.94)  | 905 (0.80)   |
| <i>kenka o uru</i>          | lit. sell a fight, 'pick a fight'  | 74 (0.71)  | 683 (0.60)   |
| <i>kenka o kau</i>          | lit. buy a fight, 'accept a challenge'                                   | 8 (0.08)   | 79 (0.07)    |

*katai* (lit. one's mouth is hard, 'able to keep a secret'). The relationship between antonymous forms may also be asymmetrical, as in *futokoro ga atatakai* (lit. one's breast pocket is warm, 'have a fat wallet'), on one hand, and *futokoro ga sabishii/samui* (lit. one's breast pocket is lonely/cold, 'be low on cash'), on the other.

**Table 5 Antonymous variants (2)**

| idioms/variants             | meaning (literal/idiomatic)   | BCCWJ     | TWC        |
|-----------------------------|---|-----------|------------|
| <i>kuchi ga karui</i>       | lit. one's mouth is light<br>1) 'talkative; loquacious'<br>2) 'a blabbermouth; unable to keep a secret' | 24 (0.23) | 115 (0.10) |
| <i>kuchi ga omoi</i>        | 1) lit. one's mouth is heavy, 'reticent/slow to speak'  | 58 (0.55) | 131 (0.12) |
| <i>kuchi ga katai</i>       | 2) lit. one's mouth is hard, 'able to keep a secret'  | 46 (0.44) | 193 (0.17) |
| <i>futokoro ga atatakai</i> | lit. one's breast pocket is warm, 'have a fat wallet'   | 13 (0.12) | 34 (0.03)  |
| <i>futokoro ga sabishii</i> | lit. one's breast pocket is lonely, 'be low on cash'  | 11 (0.10) | 46 (0.04)  |
| <i>futokoro ga samui</i>    | lit. one's breast pocket is cold, 'be low on cash'  | 10 (0.10) | 21 (0.02)  |

There are also examples of complex networks of antonymous and synonymous variants that overlap partially in terms of meaning, lexis, and grammatical function. Table 6 shows a cluster that exploits the metaphor of a curtain to indicate the beginning and end of actions and events. First, there is a set of transitive synonyms/antonyms composed of *maku o akeru* (lit. open the curtain, 'start s.t.'), on one hand, and *maku o tojiru/orosu/hiku* (lit. close/lower/draw the curtain, 'put an end to s.t.'), on the other. Secondly, there is a set that includes the intransitive antonyms *maku ga aku/oriru* (lit. the curtain opens/comes down, 's.t. starts/comes to an end') and the phrase *maku ga kitte otosareru* (lit. the curtain is cut and dropped, 's.t. starts with a bang'), which is an emphatic synonym of *maku ga aku* that uses the passive form of the verb *otosu* 'drop'.

**Table 6 Cluster of synonymous/antonymous variants**

| idioms/variants                | meaning (literal/idiomatic)                                       | BCCWJ      | TWC          |
|--------------------------------|---|------------|--------------|
| <i>maku o akeru</i>            | lit. open the curtain, ‘start s.t.’                               | 75 (0.72)  | 706 (0.62)   |
| <i>maku o tojiru</i>           | lit. close the curtain, ‘put an end to s.t.’                      | 101 (0.96) | 1,403 (1.23) |
| <i>maku o orosu</i>            | lit. lower the curtain, ‘put an end to s.t.’                      | 39 (0.37)  | 411 (0.36)   |
| <i>maku o hiku</i>             | lit. draw the curtain, ‘put an end to s.t.’                       | 14 (0.13)  | 146 (0.13)   |
| <i>maku ga aku</i>             | lit. the curtain opens, ‘s.t. starts’                             | 14 (0.13)  | 119 (0.10)   |
| <i>maku ga kitte otosareru</i> | lit. the curtain is cut and dropped,<br>‘s.t. starts with a bang’ | 19 (0.18)  | 84 (0.07)    |
| <i>maku ga oriru</i>           | lit. the curtain comes down,<br>‘s.t. comes to an end’            | 10 (0.10)  | 69 (0.06)    |

As noted in Section 1, many idioms do not have synonymous or antonymous counterparts, even if they contain a content word that has a commonly-used synonym or antonym in the Japanese lexicon (e.g. (1) *saji*/\**spūn o nageru* lit. throw the spoon (Jap.)/\*spoon (Eng. loanword), ‘give up’; *abura o uru*/\**kau* lit. sell/\*buy oil, ‘loaf; waste time’). This means that, in comparison to expressions such as *saji o nageru* and *abura o uru*, the idioms shown in Tables 1-6 have a relatively lower degree of lexico-structural frozenness. Moreover, insofar as lexico-structural frozenness is viewed as a defining property of idioms, the latter expressions are less prototypical than the former (Ishida 1998, 2004).

## 5. Idiom variants in phraseological dictionaries

### 5.1 Materials and procedures

In order to investigate the extent to which Japanese phraseological dictionaries provide a reliable account of idiom variants, the results of the corpus analysis reported in Sections 4.1 and 4.2 were compared to representations of the target idioms/variants in seven widely-used print dictionaries (12-13).<sup>6</sup>

(12) semasiological dictionaries

KIY (1982): 245 entries

HKKJ (1988): 3,500 entries

NKJ (2005): 1,563 entries

YWKJ (2007): 3,000 entries

SKKKJ (2010): 6,550 entries

(13) onomasiological dictionaries

RKJ (1992): 3,700 entries

IHKJ (1998): 1,100 entries

Print dictionaries were used because most Japanese electronic dictionaries do not contain idiom dictionaries. Of these dictionaries, only KIY (1982) and NKJ (2005) are theory- and text-based. KIY contains a relatively small number of entries in comparison to the others; however, it also includes a “Table of Commonly Used Idioms” (*Jōyō Kanyōku Ichiran*) that lists 1280 idioms. Variants listed on this table are shown in Table

<sup>6</sup> Full titles and bibliographic details of all dictionaries are given after the list of references at the end of this paper.

7 (Section 5.2), for the sake of comparison; however, these results are not included in the analysis reported in Sections 5.2-5.3.

The dictionary analysis targeted the 41 variant forms previously identified in the corpus analysis, investigating whether or not and how each form was presented in each dictionary. Four types of listings were found and coded as follows:

- ◎=listed as a headword with an entry;
- =not listed as a headword but shown somewhere in the entry of a related variant form (in a usage note and/or illustrative example, or under a marker such as *rui* ‘synonym’ or *tai* ‘antonym’);
- =listed as a headword only (no entry) and cross-referenced to a variant form with an entry;
- ×=not listed or shown in any way.

## 5.2 Dictionary analysis of synonymous variants

Results of the dictionary analysis of synonymous variants are shown in Table 7. (For the literal and idiomatic meanings of each idiom, please see Tables 1-3.)

Some dictionaries list both members of a pair of synonymous variants as individual headwords (12.5%). For example, *mi ni shimiru* (lit. permeate one’s flesh) and *honemi ni shimiru* (lit. permeate one’s bones and flesh) are both listed in three dictionaries (HKKJ/RKJ/NKJ), *shūshifu o utsu* (lit. strike [typewrite] a period (Jap.)) and *pirido o utsu* (lit. strike [typewrite] a period (Eng. loanword)) in two (YWKJ/NKJ), and *kageguchi o tataku* (lit. beat the shadow-mouth) and *kageguchi o kiku* (lit. make use of the shadow-mouth) in one (NKJ).

In a greater number of cases (36%), one expression is listed as a headword, and a synonymous variant is listed elsewhere in the entry, sometimes under the marker *rui(ku)* ‘similar (phrases)’ (YWKJ/RKJ; see also SKKKJ). *Kageguchi o tataku* (lit. beat the shadow-mouth) is listed as a headword with *kageguchi o kiku* (lit. make use of the shadow-mouth) shown as a synonymous variant in five dictionaries (YWKJ/SKKKJ/RKJ/IHKJ/KIY). This may reflect lexicographers’ intuitions that the variant with *tataku* ‘hit; beat’ is used more frequently. However, in some dictionaries *kao-iro o miru* (lit. look at s.o.’s face-colour) is listed as a headword with *kao-iro o ukagau* (lit. glance at s.o.’s face-colour) as its variant (SKKKJ/RKJ), despite the fact that the latter has a considerably higher frequency in both the BCCWJ and the TWC (Table 7).

NKJ and YWKJ are overall the most consistent of the seven dictionaries with respect to the treatment of synonymous variants, listing multiple variants either as independent headwords or as synonymous forms within an entry in just over half the cases surveyed here.

Finally, in many cases, only one of the idiom forms is listed, with no indication of an alternate form attested in corpus data (30%). HKKJ lists *pirido o utsu*, but not *shūshifu o utsu*. The only dictionary that lists *kageguchi o iu* (lit. speak/say the shadow-

Table 7 Synonymous variants in Japanese phraseological dictionaries

| synonymous variants                 | frequency |       | HK KJ | YW KJ | SKK KJ | RKJ | IH KJ | KIY (entry) | KIY (list) | NKJ |
|-------------------------------------|-----------|-------|-------|-------|--------|-----|-------|-------------|------------|-----|
|                                     | BCC WJ    | TWC   |       |       |        |     |       |             |            |     |
| <i>mi ni shimiru</i>                | 212       | 2422  | ⊙     | ⊙     | ○      | ⊙   | ×     | ×           | ⊙          | ⊙   |
| <i>honemi ni shimiru</i>            | 23        | 163   | ⊙     | ○     | ⊙      | ⊙   | ⊙     | ×           | ×          | ⊙   |
| <i>shūshifu o utsu</i>              | 164       | 1397  | ×     | ⊙     | ⊙      | ⊙   | ⊙     | ×           | ×          | ⊙   |
| <i>piriodo o utsu</i>               | 42        | 340   | ⊙     | ⊙     | ●      | ○   | ○     | ×           | ×          | ⊙   |
| <i>hadome o kakeru</i>              | 96        | 1229  | ×     | ×     | ×      | ×   | ×     | ⊙           | ⊙          | ○   |
| <i>burēki o kakeru</i>              | 61        | 614   | ⊙     | ⊙     | ⊙      | ⊙   | ×     | ⊙           | ⊙          | ⊙   |
| <i>kotoba o nigosu</i>              | 80        | 249   | ○     | ⊙     | ⊙      | ⊙   | ⊙     | ×           | ⊙          | ×   |
| <i>kuchi o nigosu</i>               | 16        | 59    | ⊙     | ○     | ×      | ×   | ×     | ×           | ×          | ×   |
| <i>kao-iro o ukagau</i>             | 98        | 778   | ⊙     | ×     | ○      | ⊙   | ×     | ×           | ×          | ×   |
| <i>kao-iro o miru</i>               | 37        | 235   | ×     | ×     | ⊙      | ⊙   | ×     | ×           | ×          | ×   |
| <i>kao-iro o yomu</i>               | 10        | 17    | ●     | ×     | ×      | ×   | ×     | ×           | ×          | ×   |
| <i>kageguchi o tataku</i>           | 59        | 257   | ×     | ⊙     | ⊙      | ⊙   | ⊙     | ○           | ×          | ⊙   |
| <i>kageguchi o iu</i>               | 20        | 197   | ×     | ○     | ×      | ×   | ×     | ×           | ×          | ×   |
| <i>kageguchi o kiku</i>             | 14        | 16    | ×     | ○     | ○      | ○   | ○     | ○           | ×          | ⊙   |
| <i>yumizu no yō ni</i>              | 43        | 371   | ×     | ×     | ×      | ×   | ○     | ×           | ⊙          | ⊙   |
| <i>yumizu no yō ni tsukau</i>       | [31]      | [187] | ⊙     | ⊙     | ⊙      | ⊙   | ×     | ×           | ⊙          | ○   |
| <i>chi no nijimu yō (na)</i>        | 28        | 255   | ○     | ⊙     | ⊙      | ⊙   | ×     | ⊙           | ⊙          | ⊙   |
| <i>chi no nijimu yō na do'ryoku</i> | [19]      | [157] | ×     | ○     | ×      | ×   | ×     | ○           | ×          | ○   |

⊙=listed as a headword/entry; ○=not listed as a headword but shown somewhere in the entry of a related variant form (in a usage note and/or illustrative example, or under a marker such as *ru* 'synonym'); ●=listed as a headword only (no entry) and cross-referenced to a variant form with an entry; ×=not listed or shown in any way.

mouth) as a synonym for *kageguchi o tataku* (lit. beat the shadow-mouth) is YWKJ, even though the former is used more often than *kageguchi o kiku* (lit. make use of the shadow-mouth). *Hadome o kakeru* (lit. put on the brake/skid (Jap.)) and *kuchi o nigosu* (lit. muddy one's mouth) also tend to be overlooked. This may be for reasons of space (cf. Miyaji 1982b: 94), or it may reflect lexicographers' intuitions of currency or stability. However, considering the frequency (Table 7) and familiarity of these variants, it is arguable that they should be included in dictionaries, in order to meet users' lookup needs as well as to provide a clear picture of the variability of individual idioms.

It should be noted that the "empty slot" idioms identified in this paper are not treated consistently. Four dictionaries list *yumizu no yō ni tsukau* (lit. use like hot and cold water) as a headword, with no indication of potential alternation with other verbs.

An exception is NKJ, which lists *yumizu no yō ni* (lit. like hot and cold water) as a headword and presents the combination with *tsukau* in a usage note and examples. On the other hand, all of the dictionaries that include *chi no nijimu yō (na)* (lit. like blood oozing) as a headword list it as such, and only a few entries (YWKJ/KIY/NKJ) show the tendency of this expression to combine with the noun *do ryoku* ‘effort’.

### 5.3 Dictionary analysis of antonymous variants

Some antonymous variants are treated systematically in Japanese idiom dictionaries. In 46% of all possible cases, both members of an antonym pair (Table 4) or at least two members of an antonym set (Tables 5 and 6) are listed as independent headwords.<sup>7</sup> This includes the pairs *ki ga mijikai/nagai* (lit. one’s *ki* [spirit/mind] is short/long) and *te o kasu/kariru* (lit. lend/borrow a hand), which each appear in four dictionaries (HKKJ/YWKJ/SKKKJ/RKJ and YWKJ/SKKKJ/RKJ/NKJ, respectively), as well as the set *kuchi ga karui/omoi/katai* (lit. one’s mouth is light/heavy/hard), which appears in five (HKKJ/YWKJ/SKKKJ/RKJ/NKJ). *Kuchi ga karui* is marked as an antonym of both *kuchi ga omoi* and *kuchi ga katai* in YWKJ and RKJ (cf. Table 5); however, most dictionaries present it as an antonym of *kuchi ga katai* only.

In fewer cases, one variant is listed as a headword, and an antonym is listed somewhere in the entry (22%). *Hagire ga ii* (lit. the teeth-sharpness is good) appears as a headword in five dictionaries (YWKJ/SKKKJ/RKJ/IHKJ/NKJ), with *hagire ga warui* (lit. the teeth-sharpness is bad) listed as its variant under the marker *tai/tsuiku* ‘antonym’ or *han(tai)* ‘opposite’.<sup>8</sup> Six dictionaries list both *futokoro ga atatakai* (lit. one’s breast pocket is warm) and *futokoro ga samui* (lit. one’s breast pocket is cold) as headwords; however, *futokoro ga sabishii* (lit. one’s breast pocket is lonely) appears as a headword in only two of these (HKKJ/NKJ). The other four list this variant as a synonym under the entry for *futokoro ga samui*. This preference may reflect lexicographers’ judgment that *samui* ‘cold’ is a better antonym for *atatakai* ‘warm’, in spite of the fact that the variant with *sabishii* ‘lonely’ is as stable as the other two (Table 5).

Some variants do not appear (6%). RKJ/NKJ have an entry for *ki ga tsuyoi* (lit. one’s *ki* [spirit/mind] is strong) but not *ki ga yowai* (lit. one’s *ki* is weak), and SKKKJ lists *kenka o uru* (lit. sell a fight) but not *kenka o kau* (lit. buy a fight). Treatment of the *maku o akeru* (lit. open the curtain) cluster in Table 6 also lacks consistency. NKJ lists all variants of this cluster, with most as independent entries. However, IHKJ and RKJ

<sup>7</sup> For the dictionary analysis, the cluster in Table 6 was split into a set of transitive antonymous variants (*maku o akeru* lit. open the curtain vs. *maku o tojiru/orosu/hiku* lit. close/lower/draw the curtain) and a set of intransitive or passive antonymous variants (*maku ga aku/maku ga kitte otosareru* lit. the curtain opens/is cut and dropped vs. *maku ga oriru* lit. the curtain comes down).

<sup>8</sup> Some of the dictionaries surveyed here make use of markers such as *tai/tsuiku* ‘antonym’ (HKKJ/RKJ/YWKJ) or *han(tai)* ‘opposite’ (IHKJ) to specify the relationship between a headword and a variant that may or may not be listed separately. NKJ subsumes synonymous/antonymous variants and semantically-related independent idioms under the marker *ruiku* ‘similar phrases’.

do not include *maku o orosu* (lit. lower the curtain) and *maku ga kitte otosareru* (lit. the curtain is cut and dropped), respectively, both of these dictionaries lack *maku o hiku* (lit. draw the curtain), and HKKJ lists none of the transitive variants (*maku o akeru/tojiru/orosu/hiku* lit. open/close/lower/draw the curtain).

At the same time, it is true that about two thirds (68%) of the antonymous variants surveyed here do appear in phraseological dictionaries. YWKJ, SKKKJ, and RKJ are the most consistent of the seven dictionaries with respect to the representation of this type of variant. About half of the pairs/sets shown in Tables 4-6 appear in these dictionaries as independent headwords with entries, and most of the rest are listed as antonymous variants within a related entry. Attention paid to this type of variant likely reflects lexicographers' intuitions that antonymous variants have the status of independent lexical items.

## 6. Conclusions

This paper has shown that corpus data is useful for the systematic extraction of idiom variants and the identification of patterns of correspondence that may be difficult to discover based on introspection alone. Future research on Japanese idioms should take advantage of corpus tools to further investigate the phenomenon of variability. In addition to the “institutionalized” variants considered here, corpus data also yield examples of ad hoc variants for the purpose of emphasis and or stylistic effect (e.g. *me kara uroko ga jū mai hodo ochita* lit. ten or so scales fell from my eyes, ‘I was completely awakened to the truth’; *ne-mimi ni ōame* lit. a heavy rain into one’s sleeping ear, ‘a huge bolt from the blue’). Future studies should explore the range and characteristics of such variation, in light of research on the creative modification of idioms in other languages (Gläser 1998; Moon 1998; Philip 2008).

The dictionary analysis reported here suggests that Japanese phraseological dictionaries tend to treat antonymous variants more systematically than synonymous variants (Sections 5.2 and 5.3). Preliminary analysis of other variant types (Section 3.1) indicates that transitive/intransitive variants are treated less systematically than antonymous variants, but more systematically than both synonymous and idiom/compound variants. Further research should extend these results and suggest a principled basis for the representation of different types of variants in dictionaries.

The results indicate a strong need for integration of the findings of past research, corpus analysis, and lexicographical practice. An important purpose of dictionaries is to provide information on usage, so phraseological dictionaries should move beyond the view that idioms are fixed expressions and present reliable descriptions of their relative variability. Since many idioms are low frequency, descriptions should take advantage of data from large-scale corpora (Moon 1998; Fellbaum et al. 2006; Ishida 2011). Future research should also address the practical issues of representing variants, such as principles for the selection of headwords, cross-referencing, and the use of markers to

indicate semantic relationships and frequency.

### 【References】

- 石田プリシラ (1998)「慣用句の変異形について—形式的固定性をめぐって—」『筑波応用言語学研究』5, pp.43-56.
- 石田プリシラ (2004)「動詞慣用句の意味的固定性を計る方法—統語的操作を手段として—」『国語学』55(4), pp.42-56.
- 伊藤眞 (1990)「慣用句とその Variation」『福岡大学人文論叢』22(2), pp.331-348.
- 伊藤眞 (1997)「第7章 言語の具象性・比喻性・受動性—日・独慣用句をめぐって—」筑波大学現代言語学研究会編『ヴォイスに関する比較言語学的研究』三修社, pp.249-297.
- 国広哲弥 (1985)「慣用句論」『日本語学』4(1), pp.4-14.
- 佐藤理史 (2007a)「基本慣用句五種対照表の作成」『情報処理学会 研究報告』178(1), pp.1-6.  
<http://kotoba.nuee.nagoya-u.ac.jp/jc2/download/kanyo/2007-03-sig1-178-1.pdf> (2014-02-28 取得)
- 佐藤理史編 (2007b)「基本慣用句五種対照表」名古屋大学大学院工学研究科電子情報システム専攻佐藤理史研究室.  
<http://kotoba.nuee.nagoya-u.ac.jp/jc2/download/kanyo/v20070528.pdf> (2014-02-28 取得)
- 宮地裕 (1982a)「慣用句解説」宮地裕編『慣用句の意味と用法』明治書院, pp.237-265.
- 宮地裕 (1982b)「動詞慣用句」『日本語教育』47(6), pp.91-102.
- 宮地裕 (1985)「慣用句の周辺—連語・ことわざ・複合語—」『日本語学』4(1), pp.62-75.
- 宮地裕 (1999)「第三章 慣用句の表現」『敬語・慣用句表現論—現代語の文法と表現の研究(二)—』明治書院, pp.213-339.
- 村木新次郎 (1985)「慣用句・機能動詞結合・自由な語結合」『日本語学』4(1), pp.15-27.
- 村木新次郎 (1991)『日本語動詞の諸相』ひつじ書房
- 森田良行 (1985)「動詞慣用句」『日本語学』4(1), pp.37-44.
- 森田良行 (1994)『動詞の意味論的文法研究』明治書院
- Fellbaum, C. et al. (2006) Corpus-based studies of German idioms and light verbs. *International Journal of Lexicography* 19(4): 349-360.
- Gläser, R. (1998) The stylistic potential of phraseological units in the light of genre analysis. In: Cowie, A.P. (ed.). *Phraseology: Theory, Analysis, and Applications*. Oxford: Oxford University Press, pp.125-143.
- Ishida, P. (2011) Corpus data and the treatment of idioms in Japanese monolingual idiom dictionaries. In: Szerszunowicz, J. et al. (eds). *International Dialogue on Phraseology. Vol. 1*. Bialystok: University of Bialystok, pp.101-127.
- Moon, R. (1998) *Fixed Expressions and Idioms: A Corpus-based Approach*. Oxford: Oxford University Press.
- Philip, G. (2008) Reassessing the canon: 'Fixed phrases' in general reference corpora. In: Granger, S; Meunier, F. (eds). *Phraseology: An Interdisciplinary Perspective*. Amsterdam: John Benjamins, pp.95-108.

### 【Dictionaries】

『意味から引ける慣用句辞典』（日本実業出版社 1998，丹野 顯編）＝IHKJ

『慣用句の意味と用法』（明治書院 1982，宮地裕編）＝KIY

『三省堂 故事ことわざ・慣用句辞典』第二版（三省堂編修所編 2010）＝SKKKJ

『日本語慣用句辞典』（東京堂出版 2005，米川明彦・大谷伊都子編）＝NKJ

『標準ことわざ慣用句辞典』（旺文社 1988，雨海博洋監修）＝HKKJ

『例解慣用句辞典』（創拓社出版 1992，井上宗雄監修）＝RKJ

『用例でわかる慣用句辞典』（学習研究社 2007，学研辞典編集部）＝YWKJ

### 【Data and tools】

NINJAL-LWP for BCCWJ (National Institute for Japanese Language and Linguistics-Lago Word Profiler for the Balanced Corpus of Contemporary Written Japanese) <<http://nlb.ninjal.ac.jp/>>

NINJAL-LWP for TWC (National Institute for Japanese Language and Linguistics-Lago Word Profiler for the Tsukuba Web Corpus) <<http://corpus.tsukuba.ac.jp/>>

### 【Acknowledgments】

This research was supported by JSPS KAKENHI Grant Number 22520388 (Grant-in-Aid for Scientific Research (C), “Stereotype Research on Idiomatic Expressions and Grammatical Phenomena”). Special thanks go to Professor Saburō Aoki for his support.